The Kalman Like Particle Filter : Optimal Estimation With Quantized Innovations/Measurements

Ravi Teja Sukhavasi and Babak Hassibi

***Abstract*— We study the problem of optimal estimation using quantized innovations, with application to distributed estimation over sensor networks. We show that the state probability density conditioned on the quantized innovations can be expressed as the sum of a Gaussian random vector and a certain truncated Gaussian vector. This structure bears close resemblance to the full information Kalman filter and so allows us to effectively combine the Kalman structure with a particle filter to recursively compute the state estimate. We call the resuting filter the Kalman like particle filter (KLPF) and observe that it delivers close to optimal performance using far fewer particles than that of a particle filter directly applied to the original problem. We also note that the conditional state density follows a, so called, generalized closed skew-normal (GCSN) distribution.**

***Index Terms*— Distributed state estimation, Sign of Innovation, Closed Skew Normal Distribution, Particle Filter, Wireless sensor network, Kalman Filter.**

I. Introduction

Recent advances in very large-scale integration and microelectromechanical system technology have led to the availability of cheap, low quality and low power consumption sensors in the market. This has generated a great deal of interest in wireless sensor networks (WSNs) due to their potential applications in several diverse fields [1]. Sensor network constraints such as limited bandwidth and power have inspired a considerable amount of research in developing energy efficient algorithms for network coverage and decentralized detection and estimation using quantized sensor observations [2]–[4].

The problem of estimation with quantized measurements is almost as old as the Kalman filter itself. An early survey on the subject can be found in [5]. However, most of the earlier techniques centered around using numerical integration methods to approximate the optimal state estimate. The advent of particle filtering [6]–[8] created a whole new set of tools to handle non-linear estimation problems. For example, [4] proposes a particle filtering solution for optimal filtering using quantized sensor measurements. But, quantizing sensor measurements can lead to large quantization noises when the observed values are large which then leads to poor estimation accuracy. In [9], this limitation is overcome by developing an elegant distributed estimation approach based on quantizing the innovation to one bit (the so called sign of innovation or SOI). In [10], this is generalized to handle multiple quantization levels. In both cases, it is assumed that the conditional state density is approximately Gaussian leading to a linear filter and a very simple characterization of its error performance. Under the Gaussian assumption, the error covariance matrix associated with the state estimation error satisfies a modified Riccati recursion of the type that appears in [11]. The only difference between this modified Riccati and the traditional one is a scaling factor λ multiplying the nonlinear term of the recursion. For the SOI Kalman filter (SOI-KF), λ is $\frac{2}{\pi}$ while [10] presents a formula for λ in the case of multiple quantization levels. Henceforth, these filters will be referred to as SOI-KF and MLQ-KF, and their associated Riccati recursions as SOI-Riccati and MLQ-Riccati respectively.

For linear time invariant dynamical systems, if the Gaussian assumption were realistic, convergence of the modified Riccati must mean the convergence of the corresponding linear filters. Using results presented in [11] one can find linear time invariant systems for which the MLQ-Riccati and SOI-Riccati converge. [12] provides examples for which the actual error performance of SOI-KF and MLQ-KF do not converge to their respective Riccatis, which warrants a closer examination of the assumption of Gaussianity. This is one of the questions addressed in this paper. We derive a novel stochastic characterization of the conditional state density (see theorem 3.1). Using this, it is straighforward to see that the conditional state density is not Gaussian, as one would expect, given the non-linear nature of quantization. In fact, it belongs to a class of distributions which we refer to as Generalized Closed Skew Normal (GCSN) distributions. A careful literature review reveals that a related observation has been made in [13]. In particular, with some effort, [13] can be used to derive theorem 3.1. Specializing this result to state space models, we develop a novel particle filtering approach to optimally estimate the state using quantized measurements/innovations. In the rest of the paper, we use the words ‘measurements’ and ‘innovations’ interchangealy since the analysis will prove that the general structure of the filter does not depend on whether sensor measurements or innovations are quantized. The proposed filter requires far fewer particles than that of a particle filter applied directly to the original problem [12], as will be shown through various simulations. We also develop a precise formulation of the conditional state density and observe that it follows what we call a generalized closed skew-normal distribution, which is very similar to those studied in [14]–[19]. Some useful properties of this distribution are also provided in the Appendix. The next section introduces the problem setup.

Ravi Teja Sukhavasi is with the department of Electrical Engineering, California Institute of Technology, Pasadena, USA teja@caltech.edu

Babak Hassibi is a faculty with the department of Electrical Engineering, California Institute of Technology, Pasadena, USA hassibi@caltech.edu

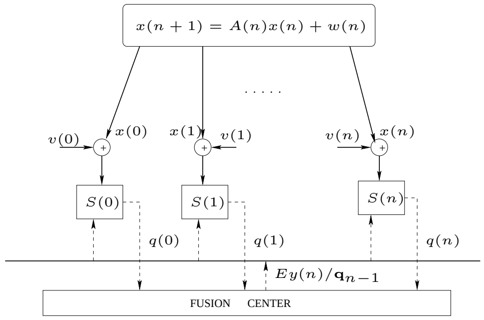


Fig. 1. WSN with a fusion center: The sensors act as data gathering devices.

II. Problem Statement and Preliminaries

The broader problem that one would like to solve can be cast as causal estimation of a random process $\{x(n)\}$ using a quantized version of a measurement process $\{y(n)\}$, where $x(n)$ and $y(n)$ can be vectors. Without any further structure, this is a difficult problem to analyze. When $\{x(n)\}$ and $\{y(n)\}$ are jointly Gaussian, we will provide a novel stochastic characterization of the probaility density of $x(n)$ causally conditioned on the quantized measurement process. We use it to propose a novel filtering technique for the above problem which reduces to an elegant particle filter when $\{x(n)\}$ has state space structure.

A. Notation

The following notation will be used in the rest of the paper.

1) If $\{u(n)\}_{n=-\infty}^{\infty}$ is a discrete time random process, $u(i\ :\ j)$ denotes the collection of random variables $(u(i), \ldots, u(j))$. $\mathcal{U}_n$ denotes the random vector $[u(0), \ldots, u(n)]^T$ and $\mathbf{u}_n$ denotes a realization of $\mathcal{U}_n$.
2) For a random vector Y, $\|Y\|^2 \triangleq E(Y-EY)(Y-EY)^T$
3) For random vectors X, Y, $\langle X, Y\rangle \triangleq E(X-EX)(Y-EY)^T$
4) For random variables $(X_1, \ldots, X_n)$, $\mathcal{L}(X_1, \ldots, X_n)$ denotes their linear span.
5) For a vector x, $x^{(i)}$ denotes its i-th component. In the context of particle filtering, x^i would denote the i-th particle.
6) $N_k(\mu, \Sigma)$ denotes a k-dim normal random variable with mean μ and covariance Σ. $N_k(a, b, \mu, \Sigma)$ denotes a k-dim normal truncated to lie in (a, b), where a, b are k-dim vectors and the truncation is component-wise.
7) $\Phi(x) = P(X \leq x)$, where $X \sim N(0, 1)$, $\Phi(a, b, \mu, \sigma) = P(X \in (a, b))$ when $X \sim N(\mu, \sigma^2)$ and in general, $\Phi(\mathcal{I}_*, \mu, \sigma) = P(X \in \mathcal{I}_*)$.
8) The notion of optimality to be used throughout the paper is mean squared error optimality.

B. Problem Formulation

Suppose $\{x(n)\}$ and $\{y(n)\}$ have the following state space structure.

$$x(n+1) = A(n)x(n) + w(n) \tag{1a}$$
$$y(n) = H(n)x(n) + v(n) \tag{1b}$$

where $x(n) \in \mathbb{R}^d$ is the state, $y(n) \in \mathbb{R}$ is the observation, and $w(n) \in \mathbb{R}^d$ and $v(n) \in \mathbb{R}$ are uncorrelated Gaussian white noises with zero means and covariances $W(n)$ and $R(n) \triangleq \sigma_v^2(n)$, respectively. The initial state, $x(0)$, of the system, is also a zero mean Gaussian with covariance $P(0)$ and is uncorrelated with both $w(n)$ and $v(n)$. As in [9], [10], we consider the sensor network configuration in which the fusion center has sufficient power to broadcast its predicted output and the corresponding error covariance to its sensors. Sensors are assumed to have limited power and hence their transmission of information should be limited. Here, we assume that the energy required for receiving messages is much less than that for transmitting.

Fig 1 outlines the overall filtering paradigm. Once a scheduling algorithm is in place, at each time instant, a sensor $S(n)$ makes a measurement $y(n)$ and computes the innovation $\tilde{e}(n) = y(n) - \hat{y}(n|n-1)$, where $\hat{y}(n|n-1) = h\hat{x}(n|n-1)$ together with the variance of the innovation $\|\tilde{e}(n)\|^2$ are received by the sensor from the fusion center, with $\hat{x}(n|n-1)$ being the one step predictor of the state. [9], [10] propose methods to quantize $\epsilon(n)$ and use the quantized innovations to update the state estimate. These filters take the following shape

$$\hat{x}(n/n) = \hat{x}(n/n-1) + L\left(q(n)\right) \frac{P(n)(H(n))^T}{H(n)P(n)(H(n))^T + \sigma_v^2}$$
$$\hat{x}(n+1/n) = A(n)\hat{x}(n/n)$$
$$P(n/n) = P(n) - \lambda \frac{P(n)(H(n))^T H(n)P(n)}{H(n)P(n)(H(n))^T + \sigma_v^2} \tag{2a}$$
$$P(n+1) \triangleq P(n+1/n) = A(n)P(n/n)(A(n))^T + W(n) \tag{2b}$$

where $q(n)$ denotes the quantized innovation while $L\left(q(n)\right)$ and the value of λ depend on the quantization scheme used. Eqs (2a) and (2b) constitute the modified Riccati recursion with parameter λ. For SIO-KF, $\lambda = \frac{2}{\pi}$ and for MLQ-KF, [10] provides a formula for λ and $L\left(q(n)\right)$. The above filter is derived based on the assumption that the conditional distribution $f\left(x(n)/\mathbf{q}_{n-1}\right)$ is Gaussian, which we will prove is generally false. [12] provides examples where the error performance of the filters in [9], [10] do not track the modified riccati recursions that they were predicted to. Instead, the optimal filter, which was approximated by a particle filter, has been observed to have an error covariance matrix that obeys the modified Riccatis. In order to approximate the optimal filter, [12] employs a very basic particle filtering algorithm which is outlined below for easy reference.

Alg 1. Particle Filter

---

1) Set $n = 0$. For $i = 1, \cdots, N$, initialize the particles, $x^i\left(0|-1\right) \sim f(x(0))$ and set $\hat{x}(0|-1) = \mathbf{0}$
2) At time n, set $q(n) = Q\left(y(n) - H(n)\hat{x}(n|n-1)\right)$, where $Q(.)$ is a quantization rule.
3) Suppose the quantized value $q(n)$ implies that $y(n) - H(n)\hat{x}(n|n-1) \in \mathcal{I}(n)$, then $v(n) + H(n)(x(n) - \hat{x}(n/n-1)) \in \mathcal{I}(n)$. The

importance weights are now calculated as follows

$$w^i(n) = \Phi(\mathcal{I}(n), H(n)(x(n) - \hat{x}(n/n-1)), \sigma_v(n))$$

4) Measurement update is given by

$$\hat{x}(n/n) = \sum_{i=1}^{N} \overline{w}^i(n) x^i(n/n-1)$$

where $\overline{w}^i(n)$ are the normalized weights, i.e., $\overline{w}^j(n) = \frac{w^j(n)}{\sum_{i=1}^{N} w^i(n)}$

5) Resample N particles with replacement accoding to,

$$\text{Prob}\left(x^i(n/n) = x^j(n/n-1)\right) = \overline{w}^j(n)$$

6) For $i = 1, \cdots, N$, predict new particles according to,

$$x^i(n+1/n) \sim f\left(x(n+1)|x^i(n|n)\right), \text{ i.e.,}$$
$$x^i(n+1/n) = A(n)x^i(n/n) + N_k(0, W(n))$$

where $W(n)$ is the covariance of the process noise at time n.

7) Set $\hat{x}(n+1|n) = A(n)\hat{x}(n|n)$. Also, set $n = n+1$ and iterate from step 2.

The particles in Alg 1 describe the conditional state density $f\left(x(n)/\mathbf{q}_n\right)$ and simulations suggest that one needs upwards of a few hundred particles to get satisfactory error performance for most systems. In the following sections, we disprove the premise behind MLQ-KF and SOI-KF and develop a novel particle filtering technique (KLPF) which converges to the optimal filter asymptotically. Simulations show that the KLPF needs far fewer particles than the particle filter of Alg 1. The difference partly lies in using particles to describe a probability density funtion with much less support than the conditional state density.

III. Full Information Vs Quantized Innovations

Suppose $\{x(n)\}$ and $\{y(n)\}$ are jointly Gaussian, then it is well known that the probability density of $x(n)$ conditioned on $\mathcal{Y}_n$ is a Gaussian with the following parameters

$$x(n)/\left(\mathcal{Y}_n = \mathbf{y}_n\right) \sim N_k(0, R_\Delta) + R_{xy}(n)(R_y(n))^{-1}\mathbf{y}_n, \tag{3}$$

$$R_x(n) \triangleq \|x(n)\|^2,\ R_y(n) \triangleq \|\mathcal{Y}_n\|^2,\ R_{xy}(n) \triangleq \langle x(n), \mathcal{Y}_n\rangle$$
$$\text{and } R_\Delta \triangleq R_x(n) - R_{xy}(n)(R_y(n))^{-1}R_{yx}(n)$$

When $\{x(n)\}$ has an underlying state space structure and $\{y(n)\}$ is a linear measurement of $\{x(n)\}$ corrupted by additive white Gaussian noise, as defined in Eq (1), it is well known that the following Riccati recursion propogates the error covariance $P(n) = R_\Delta$

$$P(n) \triangleq P(n/n-1) = A(n)P(n-1/n-1)A(n)^T + Q(n)$$
$$P(n/n) = P(n) - \frac{P(n)H(n)(H(n))^T P(n)}{H(n)P(n)(H(n))^T + \sigma_v^2} \tag{4}$$
$$P(0) = \|x(0)\|^2$$

We would like to address the problem of optimal estimation using a quantized version of the observation process. Let $Q(.)$ be a general quantization rule which generates the quantized measurement process $\{q(n)\}$, where $q(n) = Q(\mathbf{y_n})$. Note that we allow the quantization rule to depend on all the past measurements $y(i)$, $i \leq n$. This includes, as a special case, the method of quantizing the innovations first proposed in [9]. We will show that the probability density of $x(n)$ conditioned on the quantized measurements $\mathcal{Q}_n$ admits a characterization very similar to Eq (3). We state the result in the following theorem.

Theorem 3.1: The probability density of $x(n)$ conditioned on the quantized measurement $\mathcal{Q}_n$ can be expressed as

$$x(n)/\left(\mathcal{Q}_n = \mathbf{q}_n\right) \sim N_k(0, R_\Delta) + R_{xy}(n)(R_y(n))^{-1}\left(\mathcal{Y}_n/\left(\mathcal{Q}_n = \mathbf{q}_n\right)\right) \tag{5}$$

Proof: The proof is fairly straightforward. The theorem will be proved by showing that the moment generating fuction of $x(n)/\left(\mathcal{Q}_n = \mathbf{q}_n\right)$ can be seen as the product of two moment generating functions corresponding to the two random variables in Eq (5). For brevity, we will write $x(n)/\left(\mathcal{Q}_n = \mathbf{q}_n\right)$ as $x(n)/\mathbf{q}_n$.

$$f(x(n)/\mathbf{q}_n) = \int f(x(n), \mathbf{y}_n/\mathbf{q}_n) d\mathbf{y}_n$$

Noting that $f(x(n)/\mathbf{y}_n, \mathbf{q}_n) = f(x(n)/\mathbf{y}_n)$, we can write

$$Ee^{t^T x(n)}/\mathbf{q}_n = \int e^{t^T x(n)} f(x(n)/\mathbf{y}_n) f(\mathbf{y}_n/\mathbf{q}_n) dx(n) d\mathbf{y}_n$$
$$\overset{(*)}{=} e^{\frac{1}{2}t^T R_\Delta t} \underbrace{\int e^{t^T R_{xy}(n)(R_y(n))^{-1}\mathbf{y}_n} f(\mathbf{y}_n/\mathbf{q}_n) d\mathbf{y}_n}_{\triangleq \text{ mfg of } R_{xy}(n)(R_y(n))^{-1}\mathbf{y}_n/\mathbf{q}_n}$$

$$\implies M_{x(n)/\mathbf{q}_n}(t) = M_Z(t) M_{\mathbf{y}_n/\mathbf{q}_n}((R_y(n))^{-1}R_{yx}(n)t) \tag{6}$$

where $Z \sim N_k(0, R_\Delta)$. In getting $(*)$, we used the fact that $x(n)/\mathbf{y}_n \sim N_k(R_{xy}(n)(R_y(n))^{-1}\mathbf{y}_n, R_\Delta)$. For any random variable Y, it is easy to see that $M_Y(A^T t) = M_{AY}(t)$. The result is now obvious from Eq (6). ■

Comparing Eqs (3) and (5), the only difference is the measurement vector $\mathbf{y}_n$ being replaced by the random variable $\mathcal{Y}_\mathbf{q}(n) \triangleq \mathcal{Y}_n/\mathbf{q}_n$. It is easy to see that $\mathcal{Y}_\mathbf{q}(n)$ is a multivariate gaussian random variable truncated to lie in the region defined by $\mathbf{q}_n$. It is worth noting that the covariance of $x(n)/\mathbf{q}_n$, $\|x(n)/\mathbf{q}_n\|^2$, is given by

$$R_\Delta + R_{xy}(n)(R_y(n))^{-1}\|\mathcal{Y}_n/\mathbf{q}_n\|^2(R_y(n))^{-1}R_{yx}(n)$$

As the quantization scheme becomes finer (in an appropriate manner), $\mathcal{Y}_\mathbf{q}(n)$ clearly converges to $\mathcal{Y}_n$ and $x(n)/\mathbf{q}_n$ approaches a Gaussian as is well known. Using theorem 3.1, it is easy to see that $x(n)/\mathbf{q}_n$ is not Gaussian in general, contrary to the assumption made in [9], [10]. It belongs to a class of distributions, which we call the Generalized Skew Normal Distributions (GCSN), the details of which are provided in the Appendix.

We will use theorem 3.1 to propose a novel particle filtering scheme. We begin by noting that

$$Ex(n)/\mathbf{q}_n = R_{xy}(n)(R_y(n))^{-1}E\mathcal{Y}_{\mathbf{q}}(n) \tag{7}$$

The filter is implemented for the quantization scheme proposed in [9], [10]. At time n, the sensor which is scheduled to make the measurement receives a prediction of its measurement $\hat{y}(n) \triangleq \hat{y}(n/n-1)$ and the error covariance $\|y(n) - \hat{y}(n/n-1)\|^2$. The sensor, then quantizes $\frac{\tilde{e}(n)}{\|\tilde{e}(n)\|}$, where $\tilde{e}(n) \triangleq y(n) - \hat{y}(n/n-1)$, using the following quantizer and transmits the result to the fusion center.

$$Q(x) = \begin{cases} q_K & \text{if } x > r_{K-1} \\ q_i & \text{if } r_{i-1} < x \leq r_i,\ i \geq 2 \\ q_1 & \text{if } x \leq r_1 \end{cases}$$

where $(r_1, \ldots, r_{K-1})$ are the quantization levels. Using the convention, $r_0 = -\infty$ and $r_K = \infty$, we can re-write the quantization levels as $(r_0, \ldots, r_K)$. The output of the quantizer at time n will be denoted by $q(n)$ and the lower and upper limits of the interval implied by $q(n)$ will be denoted by $r_1(n)$ and $r_2(n)$ respectively. We assume that the fusion center has access to the exact value of $q(n)$ at every instant. Note that $\hat{y}(n/n-1)$ is *an* estimate of the measurement, not necessarily the optimal. Hence $y(n) - \hat{y}(n/n-1)$ is not an innovation in the true sense of the word, unless the estimator employed at the fusion center is optimal. Nevetheless, we will refer to $q(n)$ as the quantized innovation. With this setup, we will develop a filtering technique which needs far fewer particles to achieve optimal performance than the simple particle filter applied to the original problem.

IV. The Filter

We shall begin with the observation that $\mathcal{Y}_{\mathbf{q}}(n)$ is a multivariate normal distribution truncated as follows

$$\begin{aligned}\mathcal{Y}_{\mathbf{q}}(n) &= \mathcal{Y}_n/\Big(\tilde{e}(j) \in (r_1(j), r_2(j))\ \forall\ j \leq n\Big) \\ &= \mathcal{Y}_n/\Big(y(j) \in (r_1(j) + \hat{y}(j), r_2(j) + \hat{y}(j))\ \forall\ j \leq n\Big)\end{aligned}$$

Let $s_1(n) \triangleq r_1(j) + \hat{y}(j)$ and $s_2(n) \triangleq r_2(j) + \hat{y}(j)$. Then, clearly, $\mathcal{Y}_{\mathbf{q}}(n)$ is a multivariate normal with its j-th component truncated to lie in the interval $(s_1(j), s_2(j))\ \forall\ j \leq n$, i.e., $\mathcal{Y}_{\mathbf{q}}(n) \sim N_{n+1}(\mathbf{s}_{1,n}, \mathbf{s}_{2,n}, 0, R_y(n))$. From Eq (7), it is clear that the optimal state estimate can be computed by first computing the mean of the truncated normal. Before proposing the filter, we need a couple of results for

(a) relating the distributions of $\mathcal{Y}_{\mathbf{q}}(n)$ and $\mathcal{Y}_{\mathbf{q}}(n-1)$ and
(b) generating scalar truncated normal random variables.

Lemma 4.1: Let $(Z_1, Z_2, \cdots, Z_n) \sim f_n(\mathbf{z}_n)$, where

$$f_n(\mathbf{z}_n) = N_n(\mathbf{z}_n; \mathbf{s}_{1,n}, \mathbf{s}_{2,n}, 0, R_z(n)),$$

then the marginal of the first $n-1$ components is given by

$$\begin{aligned}&(Z_1, \cdots, Z_{n-1}) \sim f_n(\mathbf{z}_{n-1}) \\ &\propto \overbrace{N_{n-1}(\mathbf{z}_{n-1}; \mathbf{s}_{1,n-1}, \mathbf{s}_{2,n-1}, 0, R_z(n-1))}^{\equiv f_{n-1}(\mathbf{z}_{n-1})} \times \\ &\quad \Phi\left(s_1(n), s_2(n), Ez(n)/\mathbf{z}_{n-1}, \sqrt{R_{\Delta n}}\right)\end{aligned}$$

Also the conditional distribution of $Z_n/Z_{1:n-1}$ is given by

$$\begin{aligned}f_n(z(n)/\mathbf{z}_{n-1}) &= N\Big(z(n); s_1(n), s_2(n), Ez(n)/\mathbf{z}_{n-1}, R_{\Delta n}\Big) \\ Ez(n)/\mathbf{z}_{n-1} &= \langle z(n), \mathcal{Z}_{n-1}\rangle(R_z(n-1))^{-1}\mathbf{z}_{n-1} \\ R_{\Delta n} &\triangleq \|z(n)\|^2 - \langle z(n), \mathcal{Z}_{n-1}\rangle(R_z(n-1))^{-1}\langle \mathcal{Z}_{n-1}, z(n)\rangle\end{aligned}$$

Note that $f_n(z(n)/\mathbf{z}_{n-1})$ is a one dimensional truncated normal.

Proof: The proof is straightforward and is not presented here due to space limitations. ■

The following Lemma describes a standard technique to generate scalar truncated normal random variables.

Lemma 4.2: Let $U \sim U(0,1)$, then

$$Y = \Phi^{-1}\left((\Phi(b) - \Phi(a))\,U + \Phi(a)\right)$$

is distributed as $N(a, b, 0, 1)$

Proof: If $F(y)$ denotes the cumulative distribution function of Y, then it is well known that $F(Y) \sim U(0,1)$. We also have

$$F(y) = \begin{cases} 0 & \text{if } y < a \\ \frac{\Phi(y)-\Phi(a)}{\Phi(b)-\Phi(a)} & \text{if } a \leq y < b \\ 1 & \text{if } y \geq b \end{cases}$$

So, if $U \sim U(0,1)$, then $F^{-1}(U) \sim N(a, b, 0, 1)$ and hence the technique. Noting that $N(a, b, \mu, \sigma^2) \equiv N(\frac{a-\mu}{\sigma}, \frac{b-\mu}{\sigma}, 0, 1)$, we can generate scalar truncated normal random variables with arbitrary mean and variance. ■

We will now propose a particle filtering technique to recursively compute the state estimate.

Alg 2. Truncated Normal Particle Filter

---

1) At $n = 0$, generate $\{y^i(0)\}_{i=1}^N \sim N(s_1(0), s_2(0), 0, R_y(0))$ using the technique outlined in Lemma 4.2.
2) At time n, for each particle $\{\mathbf{y}_{n-1}^i\}$, compute the weight as
$$w^i(n) = \Phi\left(s_1(n), s_2(n), Ey(n)/\mathbf{y}_{n-1}^i, \sqrt{R_{\Delta n}}\right) \tag{8}$$
3) Generate
$$\{y^i(n)\}_{i=1}^N \sim N\left(s_1(n), s_2(n), Ey(n)/\mathbf{y}_{n-1}^i, R_{\Delta n}\right)$$
and define $\mathbf{y}_n^i = \begin{bmatrix} \mathbf{y}_{n-1}^i \\ y^i(n) \end{bmatrix}$.
4) Measurement update
$$\hat{x}(n|n) = R_{xy}(n)(R_y(n))^{-1}\frac{\sum_{i=1}^N w^i(n)\mathbf{y}_n^i}{\sum_{i=1}^N w^i(n)} \tag{9}$$
5) Resample the N particles $\{\mathbf{y}_n^i\}_{i=1}^N$ with replacement accoding to Prob $(\mathbf{y}_n^i) = \overline{w}^i(n)$ where the normalized weights are given by
$$\overline{w}^i(n) = \frac{w^j(n)}{\sum_{i=1}^N w^i(n)}$$

6) Set $\hat{x}(n+1|n) = A(n)\hat{x}(n|n)$. Also, set $n = n+1$ and iterate from step 2.

In the filter proposed above, it is important to note that the dimension of the particle $\mathbf{y}_n^i$ is n and hence increases with time. In the absence of any structure on the random processes $\{x(n)\}$ and $\{y(n)\}$, the computations involved in running the above filter will quickly become infeasible. Alg 2 is of purely theoretical interest but the technique presented above is greatly simplified when $x(n)$ has state space structure. In the following subsection, we will show how to overcome the problem of increasing particle dimension when $\{x(n)\}$ is a Gauss Markov process and $\{y(n)\}$ is a linear measurement of $\{x(n)\}$ corrupted by additive white Gaussian noise. The filter takes an elegant formulation in which the particle dimension remains fixed and is equal to the state dimension. This approach requires far fewer particles than Alg 1 and the filtering technique is quite general, in that, it can handle arbitrary number of quantization levels and arbitrary quantization intervals. We will also observe that the filter requires fewer and fewer particles as the number of quantization levels increases.

A. Exploiting State Space Structure - The Kalman Like Particle Filter

Suppose $\{x(n)\}$ and $\{y(n)\}$ have the state space structure defined in Eq (1). Then we know that $R_{xy}(n)(R_y(n))^{-1}\mathbf{y}_n^i$ is the optimal estimate of $x(n)$ given the measurement vector $\mathbf{y}_n^i$ and hence can be computed by running a Kalman filter using $\mathbf{y}_n^i$, the specific realization of the measurement vector $\mathcal{Y}_n$. Similarly $Ey(n)/\mathbf{y}_{n-1}^i$ is the optimal estimate of $y(n)$ given $\mathcal{Y}_{n-1} = \mathbf{y}_{n-1}^i$ and $R_{\Delta n}$ is the resulting error covariance. All these quantities emerge naturally in the following Kalman filtering steps.

$$x^i(0/0) = \frac{P(0)H(0)^T}{H(0)P(0)(H(0))^T + \sigma_v^2(0)} y^i(0) \tag{10a}$$

$$P(k+1) \triangleq P(k+1/k) = AP(k/k)A^T + W(k) \tag{10b}$$

$$P(k+1/k+1) = P(k) - \frac{P(k)H(k)(H(k))^T P(k)}{H(k)P(k)(H(k))^T + \sigma_v^2(k)} \tag{10c}$$

$$x^i(k+1) \triangleq x^i(k+1/k) = Ax^i(k/k) \tag{10d}$$

$$x^i(k+1/k+1) = x^i(k+1) + \tag{10e}$$

$$\frac{P(k)(H(k))^T}{H(k)P(k)(H(k))^T + \sigma_v^2(k)} (y^i(k+1) - H(k+1)x^i(k+1)) \tag{10f}$$

Now, $R_{xy}(n)(R_y(n))^{-1}\mathbf{y}_n^i$, $R_{\Delta n}$ and $Ey(n)/\mathbf{y}_{n-1}^i$ can be calculated as follows

$$R_{xy}(n)(R_y(n))^{-1}\mathbf{y}_n^i = x^i(n/n) \tag{11a}$$

$$R_{\Delta n} = H(n)P(n)(H(n))^T + \sigma_v^2(n) \tag{11b}$$

$$Ey(n)/\mathbf{y}_{n-1}^i = H(n)x^i(n) \tag{11c}$$

Now consider Eq (9), we can write it alternately as

$$\begin{aligned}\hat{x}(n|n) &= R_{xy}(n)(R_y(n))^{-1} \frac{\sum_{i=1}^N w^i(n)\mathbf{y}_n^i}{\sum_{i=1}^N w^i(n)} \\ &= \sum_{i=1}^N \overline{w}^i(n) R_{xy}(n)(R_y(n))^{-1}\mathbf{y}_n^i = \sum_{i=1}^N \overline{w}^i(n) x^i(n/n)\end{aligned} \tag{12}$$

From Eqs (11) and (12), it is easy to see that all the information in $\mathbf{y}_n^i$ is captured in $x^i(n/n)$. Hence, we only need to work with $\{x^i(n/n)\}_{i=1}^N$ at any time n. We can now desribe the new filter as follows.

Alg 3. Kalman Like Particle Filter (KLPF)

1) At $n = 0$, generate $\{y^i(0)\}_{i=1}^N \sim N(s_1(0), s_2(0), 0, R_y(0))$ using the technique outlined in Lemma 4.2. Compute $\{x^i(0)\}$ using Eq (10a) and $x^i(1/0) = A(0)x^i(0/0)$
2) Generate
$$\{y^i(n)\}_{i=1}^N \sim N\left(s_1(n), s_2(n), H(n)x^i(n), R_{\Delta n}\right)$$
Use Eq (11b) to compute $R_{\Delta n}$. Then form $\{x^i(n/n)\}$ using Eq (10).
3) At time n, for each particle $\{x^i(n)\}$, compute the weight $\{w^i(n)\}$ as
$$w^i(n) = \Phi\left(s_1(n), s_2(n), H(n)x^i(n), \sqrt{R_{\Delta n}}\right)$$
4) (*Measurement update*) Normalize the weights to get $\overline{w}^i(n) = \frac{w^j(n)}{\sum_{i=1}^N w^i(n)}$ and compute the measurement updated state estimate using Eq (12), i.e., $\hat{x}(n/n) = \sum_{i=1}^N \overline{w}^i(n)x^i(n/n)$
5) Resample the N particles $\{x^i(n/n)\}_{i=1}^N$ with replacement accoding to Prob $\left(x^i(n/n)\right) = \overline{w}^i(n)$ and compute $x^i(n+1) = A(n+1)x^i(n/n)$.
6) Set $\hat{x}(n+1|n) = A(n)\hat{x}(n|n)$. Also, set $n = n+1$ and iterate from step 2.

Remark 1: It is easy to see that the above filter can be extended to the case when the quantizer employed at the sensor is of the more general form $Q(x) = q_j$ if $x \in \mathcal{I}_j$ for $1 \le j \le K$.

From the above implementation, in terms of complexity, the KLPF is clearly equivalent to running N parallel Kalman filters. Hence, the complexity of the KLPF scales linearly in N, the number of particles. Also, it converges to the optimal filter as $N \to \infty$. But for most systems, simulations suggest that the KLPF delivers close to optimal performance for $N \le 50$. The particles in the KLPF describe the random variable $R_{xy}(n)(R_y(n))^{-1}\mathcal{Y}_{\mathbf{q}}(n)$. The support of its distribution clearly decreases with increasing quantization levels. As a result KLPF needs fewer particles as the quantization becomes finer, a property that Alg 1 does not share. This will be demonstrated through examples in Section V.

The scenario considered thus far involves one measurement per time instant. But Alg 3 can be easily extended to

TABLE I

SUMMARY OF NUMERICAL RESULTS

	Example 1		Example 2	
	SOI	2 bit	SOI	2 bit
SOI-KF	no	-	yes	-
MLQ-KF	-	no	-	yes
Alg 1	2500	10000	500	750
KLPF	500	90	25	3

handle multiple measurements from different sensors at a given time instant. Here it is assumed that the measurement noise processes are uncorrelated across different sensors. The proof is simple and is not detailed here due to space constraints.

V. SIMULATIONS

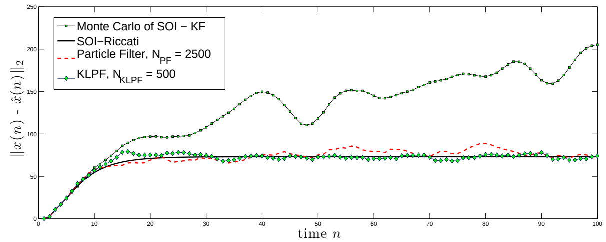
In Alg 1, the particles describe the full probability density of the state conditioned on quantized measurements. While in Alg 3, part of the information about the conditional state density is captured neatly by the Kalman filter. So, the particles describe a truncated Gaussian which has much less support than the full conditional distribution. We give two examples in this section to demonstrate the effectiveness of KLPF. The following table summarizes the highlights from the two examples

In Table I, a ‘yes’ indicates that the filter works and is close to optimal, a ‘no’ indicates that its estimation error diverges and a ‘-’ means that the quantization method does not apply to the filter. ‘SOI’ stands for ‘sign of innovation’ and ‘2-bit’ stands for a quantization rule with quantization intervals given by $(-\infty, -1.2437)$, $(-1.2437, -0.3823)$, $(-0.3823, 0.3823)$, $(0.3823, 1.2437)$ and $(1.2437, \infty)$. If the innovation falls in the interval $(-0.3823, 0.3823)$, no measurement update is done, so that 2 bits will suffice to represent the output of the above quantizer. The numbers in front of Alg 1 and KLPF denote the number of particles required to approximate the optimal filter. Clearly, KLPF requires far fewer particles than Alg 1. Also evident from Table I is the fact that KLPF needs dramatically fewer particles as the quantization becomes finer.

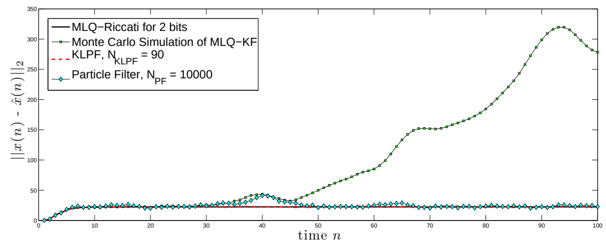
A. Example 1

We re-use this example from [12]. Consider a linear time invariant system of the form (1) with the following parameters: $A = \begin{bmatrix} 0.95 & 1 & 0 \\ 0 & 0.9 & 10 \\ 0 & 0 & 0.95 \end{bmatrix}$, $h = \begin{bmatrix} 1 & 0 & 2 \end{bmatrix}$, $W = 2\mathcal{I}_3$, $R \triangleq \sigma_v^2 = 2.5$ and $P(0) = 0.01\mathbb{I}_3$, where $\mathbb{I}_m$ denotes an $m \times m$ identity matrix. Note that A is a stable matrix. As can be seen from the plots, SOI-KF and MLQ-KF diverge but KLPF delivers optimal performance with much fewer particles than Alg 1. With the addition of just 1 bit, the required number of particles drops from 500 to 90. But with Alg 1, the number of particles goes up from 2500 to 10000.

Remark 2: Extensive simulations suggest that the modified Riccati of Eq (2) *describes the error performance of the optimal filter whereas MLQ-KF is itself clearly not the optimal filter. In Examples 1 and 2, by optimality, we meant that the mean squared errors of Alg 1 and KLPF tracked the corresponding modified Riccati recursions very closely. In fact, simulations showed that the mean squared errors of Alg 1 and KLPF saturate at this value irrespective of the number of particles used.*



(a) SOI-KF, Alg 1 and KLPF



(b) MLQ-KF, Alg 1 and KLPF

Fig. 2. *In (a), SOI-KF clearly diverges while Alg 1 and KLPF converge to the optimal filter. From (b), MLQ-KF with 4 levels of quantization also diverges while KLPF converges to the optimal filter with just 90 particles*

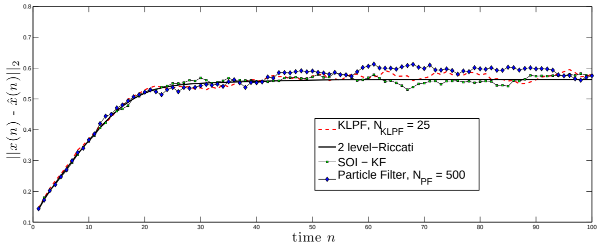
B. Example 2

A simple tracking system can be characterized by the following parameters, $A = \begin{bmatrix} 1 & \tau \\ 0 & 1 \end{bmatrix}$, $h = \begin{bmatrix} 1 & 0 \end{bmatrix}$, $W = \begin{bmatrix} \frac{\tau^4}{4} & \frac{\tau^3}{2} \\ \frac{\tau^3}{2} & \tau^2 \end{bmatrix}$, $\sigma_v^2 = 0.81$ and $P(0) = 0.01\mathbb{I}_3$ and the sampling period $\tau = 0.1$.

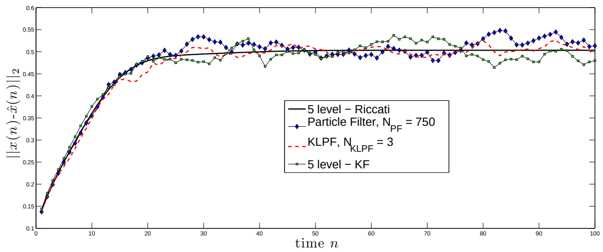
In Example 2, note that KLPF works with much fewer particles than in Example 1. One can attribute this to the much higher value of the optimal mean squared error in Example 1 than in Example 2, as can be seen from the plots.

VI. CONCLUSIONS

We propose a Kalman like particle filter (KLPF) that, in the examples studied, required moderately small number of particles and therefore can obtain close to optimal performance with a computational complexity comaparable to the conventional Kalman filter. An important open issue is to determine the number of particles necessary to closely approximate the optimal filter. As earlier observed in [12], the error covariance matrix of the optimal filter seems to follow the modified Riccati recursion introduced in [11]. Determining whether this is the case, and why, remains an interesting open question.



(a) SOI-KF, Alg 1 and KLPF



(b) MLQ-KF, Alg 1 and KLPF

Fig. 3. *Both in (a) and (b), all the filters are close to optimal. KLPF achieves optimal performance with remarkably few particles and hence has a complexity of the same order as that of SOI-KF and MLQ-KF.*

References

[1] I.F. Akyildiz, Weilian Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *Communications Magazine, IEEE*, vol. 40, no. 8, pp. 102–114, 2002.

[2] A. Krasnopeev, Jin-Jun Xiao, and Zhi-Quan Luo, "Minimum energy decentralized estimation in sensor network with correlated sensor noise," in *ICASSP*, 2005, vol. 3, pp. iii/673–iii/676.

[3] Zhi-Quan Luo and Jin-Jun Xiao, "Universal decentralized estimation in a bandwidth constrained sensor network," in *ICASSP*, 2005, vol. 4, pp. iv/829–iv/832.

[4] G Rickard Karlsson and Fredrik Gustafsson, "Particle filtering for quantized sensor information," in *13th European Signal Processing Conference, EUSIPCO*. 2005, EURASIP, Turkey.

[5] R. Curry, *Estimation and Control with Quantized Measurements*, M.I.T Press, 1970.

[6] N.J. Gordon, D.J. Salmond, and A.F.M. Smith, "Novel approach to nonlinear/non-gaussian bayesian state estimation," *Radar and Signal Processing, IEE Proceedings-F*, vol. 140, no. 2, pp. 107–113, 1993.

[7] D. Crisan and A. Doucet, "A survey of convergence results on particle filtering methods for practitioners," *IEEE Transactions on Signal Processing*, vol. 50, no. 3, pp. 736–746, 2002.

[8] M.S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.

[9] A. Ribeiro, G. B. Giannakis, and S. I. Roumeliotis, "Soi-kf: Distributed kalman filtering with low-cost communications using the sign of innovations," *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4782–4795, 2006.

[10] Keyou You, Lihua Xie, Shuli Sun, and Wendong Xiao, "Multiple-level quantized innovation kalman filter," in *17th International Federation of Automatic Control*. 2008, Seoul, Korea.

[11] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M.I. Jordan, and S.S. Sastry, "Kalman filtering with intermittent observations," in *Proceedings. 42nd IEEE Conference on Decision and Control, 2003*, 2003, vol. 1, pp. 701–708 Vol.1.

[12] Ravi Teja Sukhavasi and Babak Hassibi, "Particle filtering for quantized innovations," *ICASSP*, 2009.

[13] W.V Velde R. Curry and J. Potter, "Nonlinear estimation with quantized measurements - pcm predictive quantization, and data compression," *Information Theory, IEEE Transactions on*, vol. 16, no. 2, pp. 152–161, 1970.

[14] A Azzalini and A Dalla Valle, "The multivariate skew-normal distribution," *Biometrika*, vol. 83, no. 4, pp. 715–726, 1996.

[15] Reinaldo B. Arellano-Valle and Marc G. Genton, "On fundamental skew distributions," *J. Multivar. Anal.*, vol. 96, no. 1, pp. 93–116, 2005.

[16] Li He Marc G. Genton and Xiangwei Liu, "Moments of skew-normal random vectors and their quadratic forms," *Statistics and Probability Letters*, vol. 51, no. 4, pp. 319–325, 2001.

[17] Marc G. Genton, "Discussion of the skew normal," *Scandinavian Journal of Statistics*, vol. 32, no. 2, pp. 189–198, 2005.

[18] Marc G. Genton Philippe Naveau and Xilin Shen, "A skewed kalman filter," *J. Multivar. Anal.*, vol. 94, no. 2, pp. 382–400, 2005.

[19] Marc G. Genton, *Skew-Elliptical Distributions and their Applications - A Journey beyong Normality*, Chapman and Hall/CRC, 2004.

Appendix

A. The Generalized Closed Skew-Normal Distribution

Definition 1: For $x \in \mathbb{R}^n$, we define the generalized closed skew-normal distribution, $GCSN_{k,n}(x;\mu,\Sigma,D,s_1,s_2,\Delta)$, as follows

$$GCSN_{k,n}(x;\mu,\Sigma,D,s_1,s_2,\Delta) \triangleq N_k(x;\mu,\Sigma)\, L_{k,n}(.)$$
$$L_{k,n}(.) = \frac{\Phi_n(s_1,s_2;D(x-\mu),\Delta)}{\Phi_n(s_1,s_2;0,\Delta+D\Sigma D^T)} \tag{13}$$

$N_k(x;\mu,\Sigma)$ is a k-dim gaussian r.v and $\Phi_n(y_1,y_2;\nu,\Delta) = P(y_1^{(1)} \leq X_1 \leq y_2^{(1)},\ldots,y_1^{(n)} \leq X_n \leq y_2^{(n)})$ where $(X_1,\ldots,X_n) \sim N_n(x;\nu,\Delta)$.

The sizes of the matrices involved follow accordingly. This is a very simple generalization of the closed skew-normal (CSN) distribution defined in [19]. Infact, it reduces to the CSN if $y_1 = -\infty$ in all its components. Naturally, it inherits most of its closure properties. Skew elliptical distributions generated a lot of interest because they provide a much needed tool to handle skewness in statistical modeling and have a good number of properties in common with the standard normal distribution, such as closure under marginlization and conditioning. We will briefly discuss some of the useful properties.

Lemma 1.1: If $X \sim N_k(\mu,\Sigma)$, $Z \sim N_n(0,\Delta)$ be independent, then $X/(D(X-\mu)-Z \in (s_1,s_2)) \sim GCSN_{k,n}(\mu,\Sigma,D,s_1,s_2,\Delta)$

In the following proposition, we provide a stochastic characterization of the GCSN which can be identified as a special case of theorem 3.1 for state space systems.

Proposition 1.2: Suppose

$$V \sim N_k\left(0,\Sigma-\Sigma D^T(\Delta+D\Sigma D^T)^{-1}D\Sigma\right) \text{ and}$$
$$U \sim N_n\left(s_1,s_2,0,\Delta+D\Sigma D^T\right)$$

are independent, where $N_p(a,b,0,R)$ denotes a p-variate normal distribution truncated component-wise to lie in (a,b), where a and b are p-dim vectors specifying the boundaries of a cube in $\mathbb{R}^p$, i.e., $Z \sim N_p(a,b,0,R) \implies f_Z(z) = \frac{N_p(z;0,R)}{P(Z\in(a,b))}\mathbb{I}_{Z\in(a,b)}$.

Then, $Y = V + \Sigma D^T\left(\Delta+D\Sigma D^T\right)^{-1}U$ is distributed as $GCSN(0,\Sigma,D,\nu,\Delta)$.

The proposition will be proved using lemmas 1.3 and 1.4 which are stated below.

Lemma 1.3: If $X \sim GCSN_{k,n}(\mu,\Sigma,D,s_1,s_2,\Delta)$, then the moment generating function of X is given by

$$M_X(t) = \frac{\Phi_n(s_1,s_2;D\Sigma t,\Delta+D\Sigma D^T)}{\Phi_n(s_1,s_2;0,\Delta+D\Sigma D^T)}\exp(\mu t+\frac{1}{2}t^T\Sigma t)$$

Now, we will compute the moment generating function of a truncated multivariate normal distribution.

Lemma 1.4: Let $X \sim N_p(a,b,0,R)$, then the moment generating function of X is given by

$$M_X(t) = \frac{\Phi_n(s_1,s_2;D\Sigma t,R)}{\Phi_n(s_1,s_2;0,R)} \exp\left(\frac{t^T R t}{2}\right) \qquad (14)$$

We will now complete the proof of the proposition.

Proof: [Proof of Proposition 1.2]

Let $S = \Sigma D^T\left(\Delta + D\Sigma D^T\right)^{-1}$, then $Y = V + SU$ and $M_Y(t) = M_V(t)M_{SU}(t)$. Now, $M_{SU}(t) = M_U(S^T t)$. Plugging in the formula for the moment generating function derived in lemma 1.4, we see that $M_Y(t)$ is the moment generating function of a $GCSN_{k,n}(\mu,\Sigma,D,s_1,s_2,\Delta)$ using lemma 1.3. ■

Proposition 1.5: Let $X \sim GCSN_{k,n}(\mu,\Sigma,D,s_1,s_2,\Delta)$, $W \sim N_k(\mu_W,Q)$ be independent and define $Y = AX + W$, then $Y \sim GCSN_{k,n}(\mu_Y,\Sigma_Y,D_Y,s_1,s_2,\Delta_Y)$, where

$$\Sigma_Y = A\Sigma A^T + Q,\ D_Y = D\Sigma A^T\Sigma_Y^{-1},\ \text{and}$$
$$\Delta_Y = \Delta + D\Sigma D^T - D_Y\Sigma_Y D_Y^T,\ \mu_Y = \mu + \mu_W$$

This proof of the proposition follows directly from the following two lemmas. The first states that the GCSN is closed under full rank linear transformations while the second states its closure under addition of gaussian noise.

Lemma 1.6: If $Y \sim GCSN_{k,n}(\mu,\Sigma,D,s_1,s_2,\Delta)$ and A is a full column rank $k \times p$ $(k \geq p)$ matrix, then $AY \sim GCSN_{k,n}\left(A\mu, A\Sigma A^T, D(A^TA)^{-1}A^T, s_1, s_2, \Delta\right)$. If A is a full row rank $p \times k$ $(p \geq k)$ matrix, then $AY \sim GCSN_{k,n}(\mu_A,\Sigma_A,D_A,\nu,\Delta_A)$, where

$$\mu_A = A\mu,\ \Sigma_A = A\Sigma A^T,\ D_A = D\Sigma A^T\Sigma_A^{-1}$$
$$\Delta_A = \Delta + D\Sigma D^T - D_A\Sigma_A D_A^T$$

The following lemma states that the GCSN is closed under addition of gaussian noise.

Lemma 1.7: If $X \sim GCSN_{k,n}(\mu,\Sigma,D,s_1,s_2,\Delta)$ and $W \sim N_k(\mu_W,\Sigma_W)$ are independent, then $X + W \sim GCSN_{k,n}(\mu_{X+W},\Sigma_{X+W},D_{X+W},s_1,s_2,\Delta_{X+W})$, where $\mu_{X+W} = \mu + \mu_W$, $\Sigma_{X+W} = \Sigma + \Sigma_W$, $D_{X+W} = D\Sigma(\Sigma+\Sigma_W)^{-1}$ and $\Delta_{X+W} = \Delta + D\Sigma D^T - D_{X+W}\Sigma_{X+W}D_{X+W}^T$.

B. The Conditional State Density

We will now show that the probability density of the state conditioned on the past quantized innovations $f(x(n)/\mathbf{q}_n)$ follows a GCSN. Time update and Measurement update recursions will be provided for the parameters of the conditional density. All the parameters have very simple interpretations in terms of the statistics of the state space model. For ease of exposition, we will provide all the recursions for a time-invariant state space model. But the derivations can be trivially extended to the time varying case.

Theorem 1.8 (Time and Measurement Updates): The probability density function of $x(n)/\mathbf{q}_m$ is given by $GCSN_{k,m+1}(x(n);0,\Sigma(n/m),D(n/m),\mathbf{s}_{1,n/m},\mathbf{s}_{2,n/m},\Delta(n/m))$, for $m{=}n{-}1$, n and $\forall\ n \geq 0$. The recursions relating the parameters of the distributions of $x(n-1)/\mathcal{Q}_{n-1}$, $x(n)/\mathcal{Q}_{n-1}$ and $x(n)/\mathcal{Q}_n$ are given by

$$\Sigma(n) \triangleq \Sigma(n/n{-}1) = A\Sigma(n{-}1/n{-}1)A^T + W,$$
$$D(n) \triangleq D(n/n-1) = D(n{-}1/n{-}1)\Sigma(n{-}1/n{-}1)A^T\Sigma(n)^{-1}$$

$$\mathbf{s}_{1,n-1} \triangleq \mathbf{s}_{1,n/n-1} = \mathbf{s}_{1,n-1/n-1},$$
$$\mathbf{s}_{2,n-1} \triangleq \mathbf{s}_{2,n/n-1} = \mathbf{s}_{2,n-1/n-1},$$
$$\Delta(n) \triangleq \Delta(n/n{-}1) = \Delta(n{-}1/n{-}1)+$$
$$D(n{-}1/n{-}1)\Sigma(n{-}1/n{-}1)D(n{-}1/n{-}1) - D_n\Sigma_nD_n^T \qquad (15)$$

Eq (15) constitutes the time update. The measurement update is given by the following recursions

$$\Sigma(n/n) = \Sigma(n),\ D(n/n) = \begin{bmatrix} D(n/n{-}1) \\ H \end{bmatrix},\ \mathbf{s}_{1,n} = \begin{bmatrix} \mathbf{s}_{1,n-1} \\ s_1(n) \end{bmatrix},$$
$$\mathbf{s}_{2,n} = \begin{bmatrix} \mathbf{s}_{2,n-1} \\ s_2(n) \end{bmatrix},\ \Delta(n/n) = \begin{bmatrix} \Delta(n) & 0 \\ 0 & \sigma_v^2 \end{bmatrix} \qquad (16)$$

where $\hat{x}(n) \triangleq \hat{x}(n/n{-}1) = E(x(n)/\mathbf{q}_{n-1})$. In the rest of the appendix, we will use the notation $R(n) = \Delta(n/n) + D(n/n)\Sigma(n)D(n/n)^T \overset{(*)}{=} \Delta(n) + D(n)\Sigma(n)D(n)^T$ ((*) follows from Eq (15)).

C. Interpreting the Parameters

Theorem 1.8 outlines how the conditional state density evolves with time. But, there is little intuition in the recursions proposed there. In the following lines, we will explain the connection of each of the parameters to the state space model, so that one can afford a better insight into the overall filtering technique first proposed in [9]. We will start by noting that $\Sigma(n)$ is equal to the state covariance at time n.

Lemma 1.9: $\Sigma(n) = \|x(n)\|^2$

The next Lemma states that $R(n) = \|\mathcal{Y}_n\|^2$, i.e., $R(n)$ is the covariance of the measurement vector $[y(0),\ldots,y(n)]^T$

Lemma 1.10: $R(n) = \|\mathcal{Y}_n\|^2 = R_y(n)$

The following Lemma gives an expression for $P(\mathbf{q}_n)$

Lemma 1.11: The probability of quantized innovations is given by

$$P(\mathbf{q}_n) = \Phi_{n+1}(\mathbf{s}_{1,n},\mathbf{s}_{2,n};0,R_y(n))$$

In the following Lemma, we look at the cross covariance of $x(n)$ and $\mathcal{Y}_n$

Lemma 1.12: $\langle\mathcal{Y}_n, x(n)\rangle = D(n/n)^T\Sigma(n)$

Now, applying Proposition 1.2 on the conditional state density, we can write $x(n)/\mathcal{Q}_n{=}\mathbf{q}_n$ stochastically as

$$x(n)/\mathbf{q}_n = Z + \Sigma(n)D(n/n)^T(R(n))^{-1}\mathcal{Y}_\mathbf{q}(n),\ \text{where}$$
$$Z \sim N_k(0,\Sigma(n) - \Sigma(n)D(n/n)(R(n))^{-1}D(n/n)^T\Sigma(n))$$
$$\mathcal{Y}_\mathbf{q}(n) \sim N_{n+1}(\mathbf{s}_{1,n},\mathbf{s}_{2,n};0,R_y(n)) \qquad (17)$$

Using Lemmas 1.9, 1.10 and 1.12, one can see that

$$\Sigma(n) - \Sigma(n)D(n/n)(R(n))^{-1}D(n/n)^T\Sigma(n) =$$
$$\|x(n)\|^2 - \langle x(n),\mathcal{Y}_n\rangle\left(\|\mathcal{Y}_n\|^2\right)^{-1}\langle\mathcal{Y}_n,x(n)\rangle\ \text{and} \qquad (18)$$
$$\Sigma(n)D(n/n)^T(R(n))^{-1} = \underbrace{\langle x(n),\mathcal{Y}_n\rangle}_{\triangleq R_{xy}(n)}\underbrace{\left(\|\mathcal{Y}_n\|^2\right)^{-1}}_{(R_y(n))^{-1}} \qquad (19)$$

It is now easy to see that Eqs (18) and (19) are consistent with theorem 3.1 and hence we have

$$x(n)/\mathbf{q}_n = N_k(0,P_{kal}(n/n)) + R_{xy}(n)\left(R(n)\right)^{-1}\mathcal{Y}_\mathbf{q}(n) \qquad (20)$$

where $P_{kal}(n/n)$ is the error covariance of the Kalman filter at time step n. The methodology outlined in the appendix can be seen as an alternate way of arriving at Alg 3. However it is circumlocutious and not very intuitive. Moreover, the stochastic characterization in Eq (20) cannot be extended to the more general case of theorem 3.1 using the theory of GCSN. However, it serves as conclusive evidence that the conditional state density is not Gaussian, contrary to what is assumed in [9], [10].